Magyar nyelvű diktáló rendszer támogatása újszerű nyelvi modellek segítségével

Bánhalmi András[1], Kocsor András[1], Paczolay Dénes[1]

[1] MTA-SZTE Mesterséges Intelligencia Tanszéki Kutatócsoport, Aradi vértanúk tere 1, H-6720 Szeged, Hungary
{banhalmi, kocsor, pdenes}@inf.u-szeged.hu

Kivonat: Cikkünkben újszerű megoldásokat javasolunk a valós idejű beszédfelismeréshez szükséges nyelvi modellek területén, a felismerési pontosság és sebesség növelése érdekében. Különböző nyelvi modellek (pl. szabály alapú modellek, fonéma N-gram, szó és szócsoport N-gram modellek) párhuzamos futtatásával, illetve aggregálásával egyrészt a szó N-gram simítása, másrészt a hipotézisek számának hatékonyabb csökkentése érhető el. A szócsoport N-gramok kiértékeléséhez a szavak csoportosítását a szavak mondattani szerepét leíró MSD-kódok (Morpho Syntactic Description) [3] felhasználásával végeztük el. Az N-gram alapú statisztikai modellek hagyományos kiértékelés esetén csak az n. szó teljes felismerése után szolgáltatnak valószínűségi értékeket. Olyan eljárásokat is kidolgoztunk, amelyek használatával már az n. szó felismerésének befejezése előtt rendelkezésre állnak közelítő valószínűségi becslések.

1 Bevezetés

A számítógépek megjelenésével a dokumentumok tárolása, nyilvántartása és visszakeresése nagyságrendekkel gyorsabbá vált, azonban a szövegek és adatok bevitele még mindig túlságosan sok humán erőforrást igényel. Igaz ez az orvosi vizsgálati eredmények rögzítésére is, amely folyamat egy speciális diktáló rendszer segítségével lényegesen felgyorsítható, illetve egyszerűsíthető. Kisebb népcsoportok által beszélt, illetve speciális tulajdonságokkal rendelkező nyelvekre – mint például a magyar – egyelőre nagyon kevés diktáló szoftver látott napvilágot. Az MTA-SZTE Mesterséges Intelligencia Kutatócsoport beszédfelismerési kutatásokkal foglalkozó műhelyében kifejlesztettünk egy, a magyar nyelv automatikus felismerésére alkalmas keretrendszert, amelyre egyedi diktáló rendszerek építhetők.

A beszédfelismerő keretrendszer magját két fő modul, az akusztikai és a nyelvi modul alkotja. Az akusztikai modul egy saját implementálású Rejtett Markov Modell segítségével alkalmas a magyar nyelv beszédhangkészletének hatékony felismerésére. A beszédhangmodellek felépítése egy nagyméretű beszédadatbázis [9] alapján történt. Önmagában az akusztikai modell által szolgáltatott hipotézisek, azaz a feltételezett beszédhangsorozatok száma a bemondás hosszával exponenciálisan növekszik. A nyelvi modul feladata, hogy a lehetséges hipotézisek számát kezelhető

számúra korlátozza. A nyelvi modulba beépített tudásbázisok (nyelvtani szabályok, statisztikák) rendszerint a nagyobb hatékonyság érdekében nem a teljes beszélt nyelvet, hanem csak egy-egy szakterület speciális szóanyagát és nyelvtani szabályait modellezik. Cikkünkben egy, pajzsmirigy szcintigráfiás leletekből álló szövegkorpuszt használtunk fel a nyelvi modellek definiálására.

A nyelvi modellek területén – az ismert módszerek mellett [4] – olyan újszerű megoldásokat dolgoztunk ki és valósítottunk meg, amelyek a felismerési pontosság és sebesség növelése révén valós idejű beszédfelismerést tesznek lehetővé. A kifejlesztett nyelvi modul egyik újszerű eleme, hogy különböző nyelvi modelleket (pl. szabály alapú modellek, fonéma N-gram, szó és szócsoport N-gram modellek) képes párhuzamosan alkalmazni, különböző súlyokkal aggregálva azokat. A szócsoport N-gramok kiértékeléséhez a szavak csoportosítását a szavak mondattani szerepét leíró MSD-kódok (Morpho-Syntactic Description) [3] felhasználásával végeztük el.

Az N-gram alapú statisztikai modellek hagyományos kiértékelés esetén csak az n. szó teljes felismerése után szolgáltatnak valószínűségi értékeket. Olyan eljárásokat is kidolgoztunk, amelyek használatával már az n. szó felismerésének befejezése előtt is rendelkezésre állnak közelítő valószínűségi becslések.

2. A beszédfelismerő modul felépítése

A modern beszédfelismerő rendszerek két fő modult tartalmaznak. Az akusztikus modul a beszédhangok felismerését végzi, a nyelvi modul pedig egyfajta vezérlő szereppel rendelkezik, a nyelvileg és nyelvtanilag valószínű szerkezeteket emelve ki.

2.1 Akusztikai modellek

A közép- és nagyszótáras felismerők mindegyikének gyakorlati megvalósításakor a legkisebb egység, amelyet a rendszernek fel kell ismernie, az a beszédhang. A szavak felismerése ezen alkotóelemek felismerésén keresztül valósul meg. A téma kutatása során több különböző gépi tanuló-osztályozó algoritmus [2] hasznát javasolták a nyelvi alapegységek és az összetettebb struktúrák (szavak, mondatok) felismerésének érdekében[4]. Az ilyen algoritmusokra épülő akusztikai modelleknek a két legfontosabb ága a HMM (Rejtett Markov Modell) alapú [1] illetve a szegmens alapú megközelítés [5].

A két irányvonalban közös, hogy a mikrofonból érkező digitalizált beszédjelből kis időközönként megfelelő méretű mintát veszünk, és minden ilyen kis jeldarabból bizonyos számú jellemzőt vonunk ki [7], amelyekkel az adott jeldarabot jól tudjuk jellemezni. A beszédfelismerésben használt jellemzőkinyerő algoritmusok száma igen nagy, amelyek közül az összetettebbek a hallás és a központi idegrendszer jelfeldolgozásának vizsgálatából származó tudományos eredményeket is figyelembe veszik [7].

A HMM és a szegmens alapú megközelítések abban térnek el leginkább, hogy a HMM egységnyi jeldarabokból (adatkeretekből) építkezik, míg a szegmens alapú a feltételezett (változó hosszúságú) fonetikai szegmenseket egyben modellezi. Cikkünkben, a nyelvi modellek összehasonlításához alkalmazott beszédfelismerő kere-

trendszerként egy saját implementálású, HMM alapú akusztikus modellt használtuk fel.

2.2 Nyelvi modellek

Az akusztikus modul önmagában beszédhangsorozatokat ismer fel. Az, hogy a beszédhangsorozatok közül – egy adott természetes nyelven – melyek felelnek meg értelmes szósorozatok (mondatok) fonetikus átiratainak, azt az alkalmazott nyelvi modell dönti el. A nyelvi modell feladata tehát a lehetséges beszédhangsorozatok halmazának a szűkítése, illetve az egyes beszédhangsorozatok valószínűségének megadása.

Nyelvi modelleket általában nem a teljes természetes nyelvre készítjük, hanem annak egy szűkebb, témaorientált részén. A legalapvetőbb nyelvi modell egyetlen szótárat tartalmaz csupán, és minden szó után minden szó következhet. Ez a felismerési pontosság, és a memóriahasználat szempontjából sem hatékony megoldás. A hipotézisek számának redukálása a felismerés sebességének és pontosságának nagymérvű javulását eredményezi. Azonban a keresési tér, azaz a modell által generált nyelv redukciójának általában az szab határt, hogy a modellnek le kell fednie azokat a – természetes nyelv szerint – helyes mondatoknak a többségét is, amelyek nem álltak rendelkezésre a modell építése közben. Másik lényeges szempont, hogy a modellnek „szűknek" kell lennie, azaz a nem valószínű szósorozatokat megfelelő módon „büntetnie" kell.

A nyelvi modellek többsége 2 fő csoportból, illetve ezek kombinációiból kerül ki. Az egyik csoport a formális nyelveken alapuló szabály alapú modelleket tartalmazza. Itt a szabályok a különböző ún. szócsoportok lehetséges követési sorrendjeit írják le, ahol a szóalakok egy vagy esetleg több szócsoportba vannak besorolva.

A nyelvi modellek másik nagy csoportját a statisztikus nyelvi modellek alkotják, amelyek között a legelterjedtebb modell az ún. szó N-gram. A szó N-gram az előző (n-1) szó ismeretében megadja, hogy a rákövetkező szónak milyen a statisztikai valószínűsége; ezen statisztika alapján közelítjük a $w_1,\ldots,w_n$ szósorozat valószínűségét [4], ahol w_0, …, w_{-N+2} a mondat kezdő szimbólum:

$$P(w_1, w_2, \ldots, w_n) = \prod_{i=1}^{n} P(w_i \mid w_{i-1}, \ldots, w_{i-N+1})$$

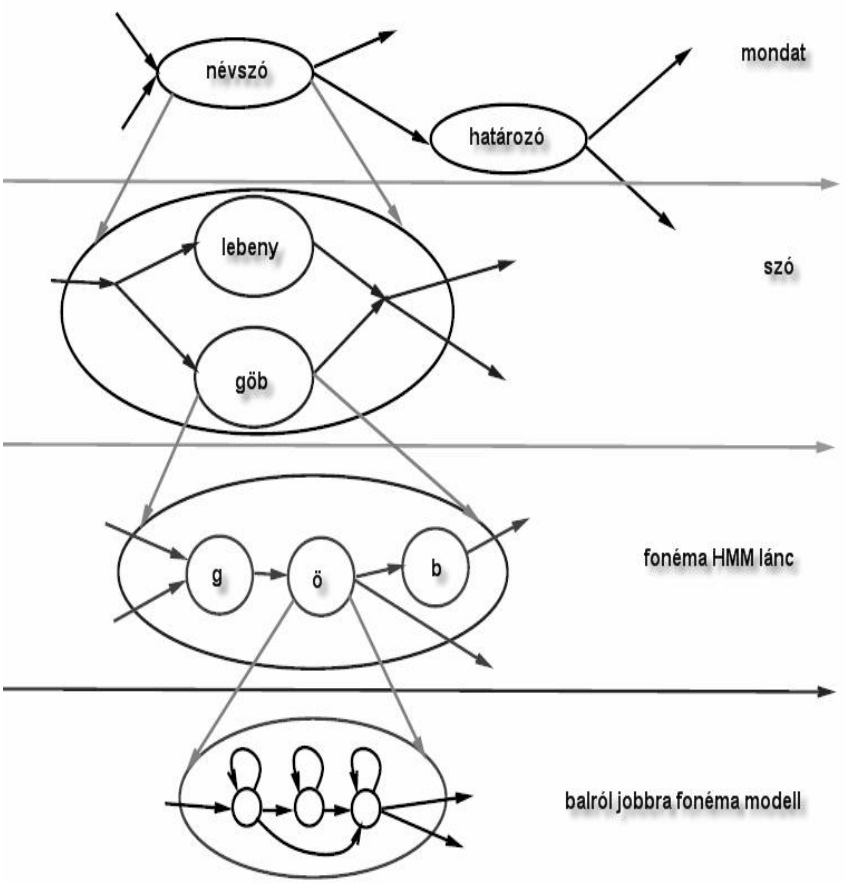
Az n növelésével egy ideig növelhető a nyelvtan megbízhatósága, de ez a memóriahasználat és a nyelvi modell tanítására használt szövegkorpusz exponenciális növekedésével jár. Emiatt a gyakorlatban 2-gram illetve 3-gram modelleket használnak. Az N-gram modellek tanítása a szövegkorpuszban található szó n-esek leszámolásával történik, azonban a rendelkezésre álló szövegkorpusz általában nem elegendő ahhoz, hogy tartalmazza kellő számban az összes bemondható helyes szó n-est, emiatt az elő nem forduló szó n-eseket rosszul modellezi a szó N-gram. A modell javítására különböző ún. N-gramsimító eljárást dolgoztak ki, amely a hiányzó vagy ritka szó n-esekhez is képes megadni valamilyen megfelelő értéket [4].

3 A nyelvi modellek kiértékelésének módszertani részletei

3.1 A kiértékeléshez használt akusztikai modul

A kiértékeléshez használt akusztikai modul egy saját HMM (Hidden Markov Model) [8] implementáció. Minden fonémához egy-egy három állapotú balról-jobbra topológiájú HMM-et rendelünk. A felismeréskor a nyelvtan által szolgáltatott beszédhangsorozat alapján HMM-ek lánca épül fel (1. ábra). Egy HMM-lánc felel meg egy hipotézisnek, azaz minden hipotézis egy-egy értelmű kapcsolatban áll egy adott beszédhangsorozattal.

1. Ábra: HMM láncot felépítő hierarchia.



A folytonos felismerő modul a hipotéziseket prioritási sorban tárolja, minden időponthoz egy-egy prioritási sor tartozik. A prioritási sort alkotó hipotéziseket pontérték szerint rangsoroljuk, amely értékeket az akusztikai modul által szolgáltatott

valószínűség, illetve a nyelvi modul szerinti pontozás határoz meg. A prioritási sor n-best vágása mellett a hipotéziseket Viterbi Beam típusú vágás segítségével is szűrjük.

A nyelvi modul adja meg, hogy az adott hipotézishez tartozó beszédhangsorozat mely fonémákkal folytatható. A nyelvi modelltől elvárjuk, hogy ezek a beszédhang kiterjesztések diszjunktak legyenek, ezáltal egy adott időponthoz tartozó prioritási sorban a hipotézisek nem ismétlődhetnek. A felismerő modul minden prioritási sornak csak az első legfeljebb *k* db hipotézisét terjeszti ki, addig, amíg be nem telik a következő időponthoz tartozó, rögzített méretű prioritási sor.

3.2 A kiértékeléshez használt nyelvi modul alapvető elemei

A saját fejlesztésű nyelvi modul egyszerre több nyelvi modell párhuzamos kiértékelésére képes. Alapvetően három független szinten képes N-gramot kiértékelni (a nyelvi modulunk értelmezi a környezetfüggetlen nyelvtani szabályokat is, de erre a jelen cikkben nem térünk ki). Az N-gramkiértékelés legalsó szintje a beszédhangok szintje (ezt jelen cikkünkben nem vizsgáljuk). A második szint a szóalakokra létrehozott N-gram. A szó N-gram – speciális esetként – tartalmaz ún. beágyazott csoportokat is, mint például a számok csoportja (pl. az 'egy', 'kettő', ..., 'kilenc', ... szavakat tartalmazó szótár szavai nem külön-külön kerülnek bele a szó N-gramba, hanem egyben, külön szabályokkal leírva). Az általunk javasolt és megvalósított 3. szintet – a bizonyos csoportok feletti N-gram modellt – a következő fejezetekben írjuk le részletesen.

3.3 A tanítás és tesztelés során használt adatbázisok

3.3.1 Az akusztikai modul tanításához használt adatbázis

Az akusztikai modul beszédhang szintű Rejtett Markov Modelljeinek tanításához a következő adatbázisokat használtuk fel:

1) MRBA beszédkorpusz: egy nagyméretű, 250 beszélő által bemondott, szegmentált adatbázis, amely összesen nagyságrendben 2000 mondatot tartalmaz. A hangadatbázis 70%-ban tartalmaz férfi, és 30%-ban női bemondást.
2) MBA beszédkorpusz: 250 beszélő által bemondott, szegmentált adatbázis, amely összesen 2500 mondatot tartalmaz. A hangadatbázis egyenlő arányban tartalmaz férfi-, illetve női bemondást. A beszélők között 50 iskoláskorú gyerek is volt.
3) OASIS-Mirigy: Az adatbázis korlátozott szókincsű és nyelvtanú mondatokat, 200 orvosi szcintigráfiás leletet tartalmaz, amely több mint 1100 mondatból, illetve kb. 11000 szóból áll.

3.3.2 A nyelvi modul tanításához és teszteléséhez felhasznált szövegkorpusz

A beszédfelismerő nyelvi modelljének létrehozásához egy pajzsmirigy-szcintigráfiás leletekből álló szövegkorpuszt használtunk. Az írásos vizsgálati anyagokat 1998 és 2004 között rögzítették. A 9231 leletet a különböző formátumokból egy közös

szöveges formátumra kellett konvertálnunk, majd többlépéses javítási folyamat következett. A vizsgálatokról készült minden egyes lelet a következő részeket tartalmazza:

a) fejléc
b) klinikai adatok
c) kérdés
d) előző vizsgálat
e) jelen vizsgálat
f) összefoglaló vélemény
g) aláírás

A szövegkorpuszból töröltük a hiányos leleteket az átvizsgálás során. A szöveges adatbázis létrehozásakor az a) és g) részek nem lettek felhasználva. A végleges adatbázis 8546 szövegből áll. A szövegekben 2500 szóalak fordul elő (számok és dátumok nélkül). Átlagosan 11 mondatot, és mondatonként 6 szót tartalmaz egy-egy lelet. A szavak számának mondatonkénti eloszlása nem normális eloszlást mutat, ami annak tudható be, hogy a közel 95000 mondat közül mindösszesen 12500 különböző mondatott tartalmazott az adatbázis.

A teszteléshez használt beszédkorpusz létrehozásakor az előbb említett szövegkorpusz bizonyos mondatai lettek beolvasva. Emiatt, a teljes átfedés elkerülése érdekében a nyelvi modell tanítását nem a teljes szövegkorpuszon, hanem annak egy részhalmazán végeztük el.

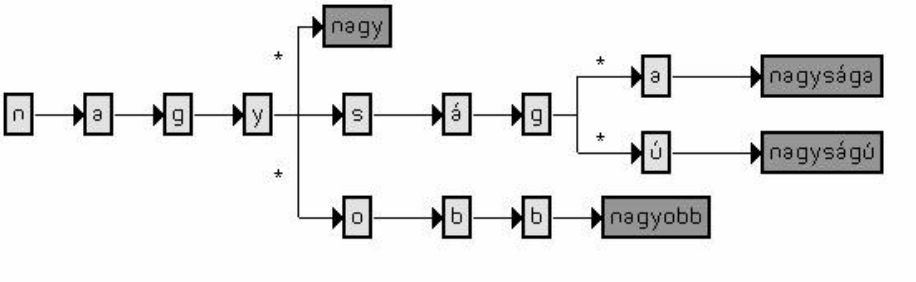
A nyelvi modellek tesztelése a Mirigy-valid nevű adatbázison történt, amely 100 szcintigráfiás orvosi lelet bemondását tartalmazza. Az összesen kb. 1000 mondat 5 beszélőtől származik.

4 Előrehozott N-gram-kiértékelés és -becslés

A szokásos N-gram kiértékelésekor egy szó n-es valószínűsége akkor derül csak ki, amikor már felismertük az n. szót is. Ha ezt a valószínűséget korábban – az n. szó vége előtt – tudnánk, akkor a hipotézisek nyelvi modul szerinti pontozása korábban megtörténne, és így csökkenthető lenne a hipotézisek száma. A megvalósítási lehetőségeket hatékonyság és memóriahasználat szempontjából vizsgáljuk meg a következő bekezdésekben.

4.1. A szótár hatékony reprezentálása

A nyelvi modellben előforduló szavakat egy tömör fastruktúrában tároljuk. A fa levelei maguk a szavak, és addig futnak közös ágon, amíg közös prefixszel rendelkeznek. Ez a tárolási módszer azért fontos, mert így minden csomópontban a lehetséges következő fonémák (többszöri előfordulás nélküli) halmaza könnyen lekérdezhető.



2. Ábra: Szavakat reprezentáló derivációs fa.

4.2 Az N-gram-kiértékelés előrehozása és becslése

A kiértékelés előrehozásának egyik kézenfekvő megoldása az, hogy a nyelvi modul a szavakat megadó derivációs fa utolsó elágazásánál már visszaadja a megfelelő N-gramértéket (lásd a 2. ábrán a csillaggal jelzett éleket). Ezt az egyszerű módszert tovább finomíthatjuk oly módon, hogy becslést adhatunk akár minden elágazásnál. Egy részfa valószínűsége felülről korlátozott, mivel a gyökeréből levezethető szavakhoz tartozó N-gram valószínűségek között van egy maximális. Ennek a meghatározására két egyszerű megoldás kínálkozik. Az egyik esetben, minden csomópontra azt is ráírjuk, hogy milyen szavak vezethetők le belőle. Ebben az esetben a csúcspont valószínűségének kiértékelésekor minden levezethető szóhoz kiszámítjuk az N-gramot, majd a maximumot adjuk vissza. Ennek a módszernek – a keresés miatt – a műveletigénye annál nagyobb, minél több szó vezethető le a csomópontból. Másik lehetőség, hogy előre kiszámítjuk és eltároljuk ezeket az értékeket minden, a becslésben részt vevő csomópontra. Mivel egy szó valószínűsége az előző (n-1)-től függ, így a csomópontokban az eltárolandó adatok száma legrosszabb esetben a szavak számának az (n-1). hatványa. Tehát ez a módszer gyors, de nagy tárigényű.

A valószínűségek előrehozott becslésével sok esetben egész hipotéziságakat tudunk levágni a felismerő modulban, tehát a nyelvi modul segítségével a valószínűtlen hipotézisek számát csökkenthetjük. Szó N-gram esetében a módszer kis-, illetve közepes méretű szótárak esetében lehet eredményes, a viszonylag nagy tárigénye miatt. A később javasolt MSD alapú csoport N-gramok esetében viszont a becsléses módszer jól alkalmazható, tárigénnyel kapcsolatos problémák nem merülnek fel.

4.3 Eredmények

A nyelvi modellek tanítása a 3.3.2 fejezetben leírtaknak megfelelően történt, a teljes szövegkorpusznak három véletlenszerűen kiválasztott részhalmazán (T1, T2, T3). Az 1. táblázat második oszlopában a hagyományos N-gram kiértékelés esetén kapott szófelismerési hibaarányt tüntettük fel. A harmadik oszlop tartalmazza az előre hozott

számítással kapott eredményeket, a 4. oszlop mutatja a relatív hibaarány csökkenését, amely azt mutatja, hogy a számítás előrehozása 12,3-18,4%-kal csökkenti a hibát.

1. Táblázat: A hagyományos (2. oszlop) és az előre hozott (3. oszlop) kiértékeléskor kapott szófelismerési hibaarányok, illetve a hibaarány csökkenése (4. oszlop)

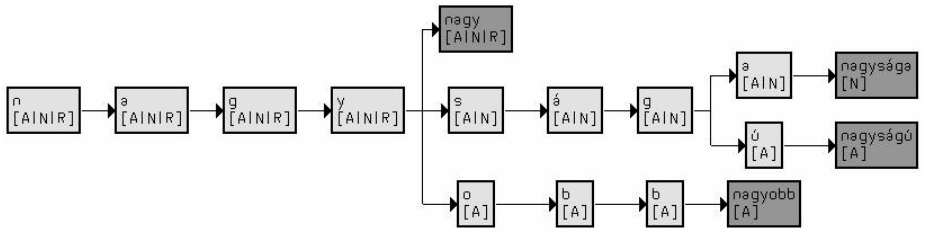
tanító korpusz	hagyományos kiértékelés	előrehozott kiértékelés	a hibaarány relatív csökkenése
T1	26.0%	22.8%	12.3%
T2	30.5%	26.3%	13.8%
T3	24,5%	20.0%	18,4%

5 MSD-kód alapú csoport N-gram

Az MSD kódolás (Morpho-Syntactic Description) minden szóhoz hozzárendel egy vagy több, változó hosszúságú karaktersorozatot, amelyek a szavak lehetséges mondattani szerepét írják le. A szavak MSD-kódjait a szegedi NLP csoport által kifejlesztett TnT alapú POS tagger programcsomag segítségével kaphatjuk meg [6]. Az MSD-kódok segítségével szabályokat hozhatunk létre a következőképpen: a szövegkorpusz mondataiban lévő szavakat lecseréljük azok mondattani szerepét leíró egyértelműsített MSD-kóddal. A cserével párhuzamosan szócsoportokat képezünk az ugyanazon MSD-kóddal rendelkező, azaz a mondatban ugyanabban a nyelvtani szerepben álló (pl. jelző, határozó) szavakból. Mivel egy-egy szónak más-más mondattani szerepe lehet a különböző mondatokban, így a csoportok nem feltétlenül diszjunktak. A folyamat végén minden mondathoz egy kódsorozatot rendelünk, ezek összessége adja meg az MSD-kódos nyelvtani szabályokat (az azonosak összevonása után). Az MSD-kód igen részletes leírást ad, így előfordulhatnak olyan csoportok, amelyekbe nagyon kevés szó esik. A szabályépítés közben, a hasonló mondattani szerepet leíró ritka kódokat érdemes összevonni, ezzel csökkenteni a csoportok számát.

MSD-kód alapú nyelvi modell nem csak konkrét, mondatokat leíró szabályhalmazként alkalmazható. Tesztjeinkben az MSD-kódolást csoport N-gram létrehozására használtuk. Az MSD-kód alapú leírásnál a szavak jelentése eltűnik, csak a szavak mondattani szerepe marad meg, így önmagában nem alkalmas nyelvi modellként. Tesztjeinkben azt vizsgáltuk, hogy alkalmazható-e, és milyen eredménnyel az MSD alapú csoport N-gram és a szó N-gram módszer kombinációja a felismerési pontosság javítására. Cikkünkben a módszerek lehetséges kombinációi közül az egyik legegyszerűbbet vizsgáljuk meg: a két modell által szolgátatott valószínűségi érték szorzata adja az aggregált értéket.

Csoport N-gramok esetében is lehetőség van előrehozott kiértékelésre. Az MSD alapú szócsoportok száma már nagyságrendekkel kisebb a szavak számánál (3. ábra), így az előre hozott becslés (4.2 fejezet) hatékonyan megvalósítható.



3. Ábra: Szavakat reprezentáló derivációs fa MSD-kódokkal kiegészítve.

A becslést a csoportok esetén hasonlóan végezhetjük el, mint szavak esetén. A tesztjeink során azt a megoldást választottunk, hogy a derivációs fa minden csomópontjához egy táblázatot rendelünk hozzá. A táblázat sorai a különböző előzményekhez tartoznak. A táblázatok kitöltéséhez minden egyes csomóponthoz meghatározzuk, hogy milyen csoportba tartozó szavak vezethetők le ebből a részfából. A csomóponthoz hozzárendelt táblázat ezután egyetlen oszlopot tartalmaz, melyben a csoport N-gram által meghatározott maximális levezethető érték szerepel (2. táblázat). A hipotézisek kiterjesztésekor a derivációs során előre haladva a levezethető csoportok száma szűkül, így a maximális érték csökken. Minden olyan csomópontban, ahol a maximális érték csökken, a nyelvi modul megadja a megfelelő arányszámot.

2. Táblázat: A 3. ábrán látható derivációs fához tartozó, MSD alapú csoport N-gram alapján számított lehetséges táblázatok. A 2. oszlopok tartalmazzák a maximummal becsült valószínűségi értékeket a fejlécében megadott csoportokra vonatkoztatva, az első oszlopokban feltüntetett előzményeket véve (- a kezdést jelenti).

	[A]
-, -	0,4
-, A	0
-, C	0
-, V	0
…	
M, N	0,25
N, A	0,20
R, A	0,16
…	

	[A|N]
-, -	0,4
-, A	1,0
-, C	0
-, V	1
…	
M, N	0,25
N, A	0,52
R, A	0,65
…	

	[N]
-, -	0,2
-, A	1,0
-, C	0
-, V	1
…	
M, N	0,005
N, A	0,52
R, A	0,65
…	

	[A|N|R]
-, -	0,4
-, A	1,0
-, C	0
-, V	1
…	
M, N	0,47
N, A	0,52
R, A	0,65
…	

5.1 Eredmények

Az MSD alapú csoport N-grammal kapcsolatos tesztjeinket az előző fejezetben már leírt tanító és tesztkorpuszokon végeztük. A 3. táblázat 2. oszlopa a csak szó N-gram használatakor kapott szófelismerési hibaarányt tartalmazza (előre hozott számítást

alkalmazva). Az MSD alapú csoport N-gram előre hozott számításával kapott teszteredményeket a 3. oszlop mutatja. Az utolsó oszlopban lévő eredményeket a csoport N-gram előrehozott becslésével kaptuk. A táblázatból kiolvasható, hogy minden esetben javított a csoport N-gram modellhez való hozzávétele, illetve a becslés. A 4. táblázatban a szófelismerési hiba relatív csökkenését foglaltuk össze.

3. Táblázat: Szófelismerési hibaarány MSD-kód alapú csoport N-gram használata nélkül (2. oszlop), csoport N-gram előre hozott számításával (3. oszlop), valamint folyamatos becslés esetén (4. oszlop).

tanító korpusz	csoport N-gram nélkül	előre hozott csoport N-grammal	becsült csoport N-grammal
T1	22.8%	21.4%	18.8%
T2	26.3%	21.2%	19.0%
T3	20.0%	17.5%	15.7%

4. Táblázat: Szófelismerési hibaarány relatív csökkenése csoport N-gram előre hozott számításakor (2. oszlop), valamint folyamatos becslés esetén (3. oszlop).

tanító korpusz	előre hozott csoport N-grammal	becsült csoport N-grammal
T1	6.1%	17.5%
T2	19,4%	27.8%
T3	12.5%	21.5%

6. Összefoglalás

Cikkünkben olyan módszereket adtunk meg, amelyek segítségével a hagyományos szó N-gram alapú nyelvtanokkal elérhető szófelismerési hibaarány csökkenthető. Tesztekkel igazoltuk, hogy a szó N-gram kiértékelésének előrehozásával a felismerés pontossága növekszik. A hagyományos szó N-gram rosszul kezeli azokat a bemondásokat, amelyek nem voltak benne a tanításához használt korpuszban. Ennek kiküszöbölésére cikkünkben a szó N-gram és az MSD típusú csoport N-gramértékeinek aggregációját javasoljuk. A teszteredményeink alapján a felismerés hibája nagymértékben csökken az aggregációs technika használatakor. Összességében az aggregációs technikával és a csoport N-gram, valamint a szó N-gram előrehozott számításával akár több mint 30%-os szófelismerési hibaarány-csökkenés érhető el.

Bibliográfia

1. C. Becchetti, L. P. Ricotti: Speech Recognition, John Wiley & Sons LTD, Chichester, England (2000)

2. R. O. Duda, P. E. Hart, D. G. Stork: Pattern Classification, Wiley (2001)

3. Erjavec, T., Monachini, M., (ed.): Specification and Notation for Lexicon Encoding. Copernicus Project 106 "MULTEX-EAST", Work Package 1 - Task 1.1, Deliverable D1.1F, (1997)

4. X. Huang, A. Acero, H. Hon: Spoken Language Processing, Prentice Hall, New Jersey (2001)

5. A. Kocsor, A. Kuba, L. Tóth , M. Jelasity, L. Felföldi, T. Gyimóthy, J. Csirik: A Segment-Based Statistical Speech Recognition System for Isolated/Continuous Number Recognition, Proceedings of the FUSST'99, Aug. 19-21, Sagadi, Estonia, 201-211, (1999)

6. Kuba A., Hócza A., Csirik J.: POS Tagging of Hungarian with Combined Statistical and Rule-based Methods in Proc. of the Seventh International Conference on Text, Speech and Dialogue (TSD 2004), Brno, Czech Republic 8-11 September, pp. 113-121 (2004)

7. L. R. Rabiner, R. W. Schafer: Digital Processing of Speech Signals, Prentice-Hall, Englewood (1978)

8. V. N. Vapnik: Statistical Learning Theory, Wiley (1998)

9. Vicsi Klára, Kocsor András, Teleki Csaba, Tóth László: Beszédadatbázis irodai számítógép-felhasználói környezetben, II. Magyar Számítógépes Nyelvészeti Konferencia, (2004)